PEDAGOGIEK
www.pedagogiek-online.nl
Uitgave: Amsterdam University Press

Beter leren rekenen gaat samen met grotere zekerheid, beter leren spellen met meer twijfel

Wim van Bon en Cecile Kuijpers

PED 36 (1): 49–70
DOI: 10.5117/PED2016.1.VBON

Abstract

Increasing competence in arithmetics leads to greater confidence, but with improvement in spelling doubt increases

To explore the relation between academic confidence and ability, an arithmetic and a spelling test were administered to 166 students in grades 4 and 5. For each item, students indicated whether they were confident about their answer. The agreement ('calibration') between confidence and test performance is limited. Overestimation of performance exceeds underestimation. Confidence is not a general characteristic of a student, but is dependent upon school domain and ability in that domain. Overestimation of arithmetic performance hardly differs between grades, but overestimation of spelling turns into underestimation. Apparently, the increase in ability leads to an increase in confidence in case of arithmetic, but turns into 'suspicion' in the spelling domain. Boys are more confident than girls, even if the answer is wrong. Girls excel in the identification of wrong answers.

Keywords: spelling, arithmetic, competence, self-assessment, confidence, calibration

Inleiding

Om vast te stellen hoe goed een leerling is in een schoolvak zoals lezen, spellen of rekenen kun je hem of haar een stukje laten lezen, wat laten schrijven of sommen laten maken. Als de leestekst, het dictee of de sommen goed zijn gekozen dan geeft de verhouding tussen wat goed en fout gelezen, geschreven of gerekend is belangrijke informatie over de voortgang van het leren, wellicht de meest belangrijke.

Maar het beeld van de competentie van de leerling is daarmee niet compleet. Het is bijvoorbeeld niet onwaarschijnlijk dat leerlingen die dezelfde dicteewoorden goed schrijven, verschillen in de zekerheid waarmee ze dat doen. Dat is niet alleen interessant, het is ook van praktisch belang. Een onzekere en twijfelende leerling zou woorden een volgende keer anders – fout in plaats van goed, of andersom – kunnen schrijven, terwijl de zekerder leerling de woorden een volgende keer op dezelfde manier – en dus weer goed of fout – zal schrijven. De zekerder leerlingen geven dan een stabieler foutenpatronen te zien, en daarmee een betere basis voor bijsturing van hun onderwijs, bijvoorbeeld bij *remedial teaching*, dan de onzekerder leerling. Dat geldt uiteraard niet alleen voor spellen, rekenen en lezen, maar voor alle leren.

Het doel van het onderzoek waarvan hier verslag wordt gedaan is voor twee centrale vakken van het primaire onderwijs, rekenen en spellen, te verkennen hoe de zekerheid over de antwoorden op de spelling- en rekenopgaven zich verhoudt tot de feitelijke scores. Dat gebeurt bij leerlingen van de Groepen 6 en 7, bij wie de ervaringen en meningen inzake hun eigen schoolprestaties al enigszins gevormd en gestabiliseerd kunnen zijn. De accuratesse van de zekerheidsoordelen wordt bestudeerd en hoe zij samenhangt met toenemende spelling- resp. rekenvaardigheid. Onderzoek bij zo verschillende schoolvakken maakt het mogelijk na te gaan of de zekerheid en de accuratesse persoonsgebonden zijn of ook bepaald worden door het vaardigheidsdomein. Nagegaan wordt of jongens in deze zekerheid verschillen van meisjes.

Het is van belang te weten of het juist is wat je denkt te weten en of je goed doet wat je denkt te kunnen. Dat is bijvoorbeeld nodig om te beslissen of je het werk dat je hebt gedaan moet controleren en of je je nog verder moet bekwamen op een bepaald gebied van kennis of kunde. Onderzoek naar zulke zekerheidsoordelen – gewoonlijk beschouwd als een aspect van de metacognitie (bijv. Roebers, Cimeli, Röthlisberger, & Neuenschwander, 2012; Schneider, 2001; maar zie Stankov & Lee,2008) – is veelal gedaan bij jongeren en volwassenen, vooral op het gebied van het ooggetuigenverslag (bijv. Bonham & Gonzalez-Vallejo, 2009) en dat van de kennisverwerving in het hoger onderwijs (bijv. Marsh & O'Mara, 2008). In de uitkomsten van dat onderzoek tekenen zich wel tendensen af, maar er is nog veel ongewis, zoals we zullen zien.

Fischhoff, Slovic, en Lichtenstein (1977) behoorden tot de eersten die onderzoek deden naar zekerheidsoordelen en constateerden dat ook bij volwassenen deze oordelen "te vaak" niet *gecalibreerd* (evenwichtig) zijn,

dus niet in overeenstemming met de feiten. Onderzoek heeft weliswaar laten zien (Krebs & Roebers, 2012; Lundeberg, Fox, Brown, & Elbedour, 2000; Roebers, 2002) dat de *miscalibrati*e, vooral in de vorm van overschatting, afneemt met toenemende competentie op het getoetste vakgebied en met het ouder worden, maar dat ook volwassenen er niet vrij van zijn. De mate van zekerheid hangt dus wel samen met de feitelijke kwaliteit van de prestaties, maar slechts in beperkte mate: "Those who know more, do know (slightly) more about how much they know" (Pallier et al., 2002, p. 293).

Overschatting wordt het vaakst aangetroffen bij betrekkelijk moeilijke taken en opdrachten en bij gemakkelijker taken vaker evenwicht of onderschatting (bijv. Lundeberg et al., 2000), het zogenaamde 'hard-easy effect'. Dat zou volgens Juslin, Winman, en Olsson (2000) een effect van statistische aard kunnen zijn. Controlerend voor mogelijke statistische oorzaken vonden Lundeberg et al. (2000) en Jonson en Allwood (2003) in hun data echter geen bevestiging voor dit vermoeden.

Een verklaring voor geringere overschatting bij grotere competentie op een vakgebied en bij relatief gemakkelijke opgaven wordt gesuggereerd door Kruger en Dunning (1999). Uitgaande van het acceptabele uitgangspunt dat wie een taak uit te voeren heeft zal kiezen voor de in zijn ogen meest redelijke optie, stellen zij dat onvolleerde personen geplaagd worden door een dubbele 'last': Incompetentie leidt niet alleen tot een onjuiste oplossing of conclusie, maar belet ook die fout te onderkennen, met als gevolg dat de vaardigheid of capaciteit overschat zal worden. Hoe groter de incompetentie, hoe groter de overschatting van de toetsscores. Onkunde en onwetendheid beperken aldus de metacognitieve vaardigheden nodig voor het evalueren van de taakuitvoering en leiden zo tot prestatie-overschatting. Studenten die op de experimentele taken het slechtst presteerden overschatten hun prestaties het meest en verbeterden hun zelf-evaluatie het minst nadat ze het werk van anderen hadden beoordeeld, wat in overeenstemming is met de gedachte dat gebrekkige metacognitie medieert in de totstandkoming van zelfoverschatting. Opmerkelijk is dat bij de best presterenden consequent enige onderschatting werd aangetroffen. Kruger en Dunning verklaren dat als het effect van de (onjuiste) verwachting van deze proefpersonen dat medestudenten wel even goed gepresteerd zouden hebben als zijzelf.

Ook gaat het kennelijk om een "… confidence trait that is generalizable across many domains of behavior" (Pallier et al., 2002, p. 295), want er is een sterke samenhang in de zekerheid die iemand vertoont op verschillende domeinen van kennis en vaardigheid (Jonson & Allwood, 2003; Roe-

bers, 2002; Schraw, Dunkle, Bendixen, & DeBacker Roedel, 1995), een trek die bovendien tamelijk stabiel is over de tijd (Jonson & Allwood, 2003). Vanwege de stabiliteit over taken en tijd concludeerden Kröner en Bierman (2007) dat zekerheidsoordelen niet afgestemd worden op kenmerken van de te vervullen opdrachten. Ook Pieschl (2009) stelt dat zulke oordelen blijkens hun stabiliteit vooral berusten op eerdere ervaring, niet aangepast worden aan de opgave van het moment en niet moeten worden opgevat als weergave van *lopende* ('*ongoing*') metacognitieve processen. Dat standpunt moet echter wel genuanceerd worden. Enige domeinspecifieke zekerheid is niet uitgesloten, want de correlatie tussen de zekerheidsscores op verschillende toetsgebieden is doorgaans niet maximaal en een algemene zekerheidsfactor verklaart de betrouwbare variantie in de zekerheidsscores niet volledig, zoals bijvoorbeeld de resultaten van Schraw et al. (1995) en Stankov en Lee (2008) laten zien.

Zoals gebruikelijk in gedragswetenschappelijk onderzoek, is er ook aandacht geweest voor eventuele geslachtsgebonden verschillen. Stankov (1999), bijvoorbeeld, kwam tot de slotsom dat mannen en vrouwen niet verschillen in de realiteitswaarde van hun zekerheidsoordelen. Lundeberg et al. (2000) constateerden geen geslachtsverschil in zekerheid, ook niet in interactie met etniciteit, maar herinneren aan een eerdere bevinding dat vrouwen zich meer bewust waren van foute antwoorden (Lundeberg et al., 2000). Ook Jonson en Allwood (2003) vonden geen evidentie ten gunste van de volgens hen gangbare overtuiging dat mannen meer geneigd zijn tot zelfoverschatting dan vrouwen. Kruger en Dunning (1999) vonden evenmin zo'n geslachtsverschil. Stankov en Lee (2008) constateerden daarentegen meer zelfoverschatting bij mannen dan bij vrouwen, echter in interactie met etniciteit. Onze slotsom is voorlopig dat, als er al een verschil in zelfoverschatting is, die groter is bij mannen.

Over de zekerheid van leerlingen in het leeftijdsbereik en op de leerstofgebieden van de basisschool is weinig bekend. Tekstbegrip is enige malen onderwerp geweest van onderzoek. Kasperski en Katzir (2013) concluderen uit onderzoek dat anderen vooral in het voortgezet onderwijs hadden gedaan, dat de zekerheidsoordelen bij toetsen voor begrijpend lezen beduidend gecorreleerd zijn met de feitelijke scores, hoewel deze samenhang van onderzoek tot onderzoek verschilt. De covariatie is het grootst als de moeilijkheid van de tekst overeenkomt met de leesvaardigheid van de leerlingen, dus niet te hoog of te laag is. De zelf-inschatting van zwakke begrijpend-lezers was echter minder accuraat, dus minder gecalibreerd, dan die van goede lezers; zij hadden meer de neiging tot overschatting. Dit

laatste werd bevestigd in hun eigen onderzoek bij vierde-klassers: de inschatting door de kinderen in het bovenste kwartiel van de leesscoreverdeling was tamelijk accuraat, terwijl de kinderen in het laagste kwartiel – maar ook die in de midden-kwartielen – hun prestaties aanmerkelijk overschatten.

Er is slechts weinig onderzoek gepubliceerd naar de zekerheid over de antwoorden bij de basisschoolvakken die we tot doel van ons onderzoek gekozen hebben, spellen en rekenen. Vroeg onderzoek op het terrein van de *spelling* is dat van Adams en Adams (1960) die meer overschatting dan onderschatting aantroffen bij volwassenen aan wie zij vroegen van een groot aantal woorden de goede schrijfwijze te kiezen uit telkens vier plausibele spellingen (*recognitie*) of ze te schrijven (*reproductie*), en per woord aan te geven hoe zeker ze waren het goede antwoord gegeven te hebben. Over het geheel genomen overtrof de zekerheid de correctheid, met een grotere overschatting bij reproductie dan bij recognitie.

In een onderzoek naar het effect van enkele manieren van nakijken bij het spellen vroegen Block en Peskowitz (1990) 9- tot 11-jarigen voor en nadat zij gedicteerde woorden opschreven of ze het woord goed of fout zouden schrijven resp. geschreven hadden. In de conditie waarin de leerlingen hun werk niet nakeken en die het meest representatief zal zijn geweest voor de 'natuurlijke' situatie, correleerde de inschatting vooraf .67 en de inschatting achteraf .75 met de feitelijke kwaliteit van de spelling. Door het doen wordt de evaluatie blijkbaar adequater.

Paffen en Bosman (2005) lieten, ook in het kader van een trainingsonderzoek, basisschoolleerlingen zekerheidsoordelen geven over dicteewoorden die ze gehoord maar nog niet geschreven hadden. Ze stelden bij de voormeting van dat onderzoek vast dat de basisschoolleerlingen (Groep 5) 85% van de dicteewoorden dachten goed te zullen schrijven, hoewel ze toch een derde van die woorden fout schreven. Van de 15% die ze dachten fout te gaan schrijven, schreven ze ongeveer een vijfde toch goed. Onderschatting komt dus ook voor, maar minder vaak dan overschatting. De mate van overeenstemming is echter geen stabiel (persoonlijkheids)kenmerk, want hij bleek te beïnvloeden door oefening. Verbetering van het 'spellingbewustzijn' lijkt bovendien samen te gaan met verbetering van de spellingvaardigheid, wat een praktisch argument is om de aard van deze calibratie en de factoren die daarin bepalend zijn te exploreren.

Ook het onderzoek van Roebers et al. (2012) bij tweede-klassers laat overschatting van de spellingprestaties zien. Er is nauwelijks verband tussen de spellingscores van de leerlingen en hun gemiddelde zekerheidscores, en ook niet tussen itemmoeilijkheid en zekerheid. Wel waren de leer-

lingen gemiddeld zekerder over goede spellingen dan over foute spellingen, een enigszins paradoxaal resultaat, dat Roebers et al., onverklaard laten. Als dit resultaat betrouwbaar was, moeten de leerlingen in enige mate beseft hebben of hun woordspelling adequaat verlopen is. Mogelijk is dat besef zo woordafhankelijk dat enige systematiek niet herkenbaar is als de data over personen (moeilijkheidsgraad) of woorden (spellingtoetsscore) worden samengevat. Deze uitkomst wijst er wel op dat de zekerheidsoordelen van tweede-klassers slechts een beperkte bruikbaarheid hebben en dat we ons verkennend onderzoek beter bij oudere leerlingen kunnen beginnen.

Uit de samenhang van de feitelijke scores met de zekerheidsoordelen in het weinige onderzoek naar het *rekenen* mag men constateren dat er een beduidende, maar beslist geen perfecte overeenkomst tussen beide is. Desoete (2009) vond correlaties van .53 met de oordelen voor en .57 met die na het maken van de opgaven. Ook Marsh, Roche, Pajares en Miller (1997) vonden – bij high school-studenten – beduidende, maar niet-perfecte samenhang tussen de zekerheid dat zij specifieke toetsitems konden oplossen en de feitelijke scores op die items. Uit beide onderzoeksverslagen is niet af te leiden hoeveel van de discrepantie bestaat uit overschatting en hoeveel uit onderschatting. Boekaarts en Rozendaal (2010) rapporteerden meer zelfoverschatting bij jongens dan bij meisjes in de vijfde klas.

Het uitgangspunt voor ons onderzoek, de aanname dat rekenen en spellen onderscheiden, d.w.z. niet perfect gecorreleerde vaardigheden zijn, lijkt terecht. Scores op spelling- en rekentoetsen correleren op de basisschool weliswaar aanzienlijk, maar toch verre van maximaal. Bij de Cito-eindtoetsen van 2009 en 2010 bijvoorbeeld was de samenhang .56 en .52 (Van Boxtel, Engelen, & De Wijs, 2011). Roebers et al. (2012) vonden bij tweedeklassers nog lagere correlaties, rond .30. Deze beperkte correlatie is geen wonder, gegeven het verschil in de aard en de gebruiksdoelen van deze twee schoolvakken. Bij de vergelijkingen die we in dit onderzoek maken spellen en rekenen zal het dus mogelijk zijn naast overeenkomsten (door persoonsgebonden zekerheid) ook verschillen (door taakdomein-gebonden zekerheid) te vinden. Als de zekerheid niet alleen door de persoon maar ook door de taak wordt bepaald, zal de overeenstemming tussen prestatie en zekerheidsoordeel bij rekenen vermoedelijk groter zijn dan bij spellen. Bij rekenen is immers de mogelijkheid tot verificatie groter, met name omdat bij het spellen een regelgeleide aanpak doorkruist wordt door woordspecifieke voorschriften.

Als we in ons onderzoek de calibratie bij spellen en rekenen verkennen, verwachten we – gegeven de vermelde eerdere bevindingen – dat die bij beide niet perfect zal zijn en vooral dat er meer overschatting zal zijn dan onderschatting. Te verwachten is ook dat bij toenemende competentie, en dus in opeenvolgende groepen van het basisonderwijs, de overeenstemming tussen feitelijke prestaties en de zelfbeoordeling zal verbeteren, vermoedelijk zonder vrijwel perfect te worden.

Als we in de zekerheidsoordelen een verschil vinden tussen jongens en meisjes, dan zal dat vooral bestaan uit zelfoverschatting door de jongens. De onderzoeksuitkomsten bij volwassenen wijzen daar op en jongens uiten doorgaans meer vertrouwen over hun kunnen dan meisjes (Boekaerts & Rozendaal, 2010; Feingold, 1994; Helmke & Van Aken, 1995; Job & Klassen, 2012). Er zijn echter ook interacties van geslacht en schoolvak te verwachten. Dat jongens doorgaans beter zijn dan meisjes in rekenen (bijv. Boekaerts & Rozendaal, 2010; Van Boxtel et al., 2011), terwijl meisjes vaak (iets) beter spellen dan jongens (bijv. Helmke & Van Aken, 1995; Hemker, Kuhlemeier, & Van Weerden, 2010; Van Boxtel et al., 2011) zou moeten leiden tot een interactie van geslacht en schoolvak, niet alleen in prestatiescores, maar mogelijk ook in de zekerheidsoordelen. Een geslachtsgebonden verschil in zekerheid en zelfoverschatting wordt overigens niet overal en altijd gevonden (Jonsson & Allwood, 2002; Lundeberg et al., 2000; Lundeberg, Fox, & Puncochar, 1994; Roebers, 2002). Maar wellicht – zo laat bijvoorbeeld het onderzoek van Jacobs, Lanza, Osgood, Eccles, en Wigfield (2002) zien – zijn er meer factoren in het spel. Zij constateerden bijvoorbeeld dat jongens op de basisschool met meer zelfvertrouwen aan rekenen beginnen dan meisjes, maar dat hun zelfvertrouwen sneller afneemt dan dat van de meisjes, zodat aan het begin van het vervolgonderwijs het verschil verdwenen is. Op het vlak van taal is er niet zo'n beginverschil, maar is er zo'n groot verschil in daling dat op het eind van de lagere school het zelfvertrouwen van de jongens op dat gebied 'dramatisch' lager is dan dat van de meisjes.

Samenvattend: Allereerst stellen we vast hoe de verhouding is tussen de reken- en de spellingscores en hoe zij samenhangen met jaarklas (Groepen 6 en 7) en geslacht. De centrale vraagstelling die daarop aansluit is naar de weerspiegeling van dit patroon in de zekerheidsoordelen. Is de zelfbeoordeling accuraat of is er bij deze leerlingen en bij deze vakken sprake van zelfoverschatting? Zijn de oordelen over goed gemaakte opgaven accurater dan over fout gemaakte opgaven? Verschillen de geslachten en de jaarklassen in deze aspecten van calibratie? Ook komt de vraag aan de orde of rekenen en spellen beschouwd kunnen worden als competenties die

zich min of meer onafhankelijk van elkaar ontwikkelen en of de zekerheid op deze twee domeinen, rekenen en spellen, aspecten zijn van een en dezelfde algemene zekerheid.

Methode

Proefpersonen

De gegevens zijn afkomstig van 166 leerlingen van de Groepen 6 en 7 van drie scholen voor regulier basisonderwijs in en nabij Nijmegen. Het onderzoek werd uitgevoerd bij 93 leerlingen uit Groep 6 (48 jongens, 45 meisjes), en 73 leerlingen uit Groep 7 (36 jongens, 37 meisjes). De gemiddelde leeftijd van de leerlingen in Groep 6 was 9.62 jaar (*SD* = 0.62) en in Groep 7 10.64 jaar (*SD* = 0.63). Het aantal doubleurs was respectievelijk 8 en 7. De leerlingen kregen in de schoolweken gemiddeld vier tot viereneenhalf uur rekenles met de methodes Pluspunt of De Wereld in Getallen en een tot anderhalf uur spellingonderwijs met Woordspel of Taal Actief.

Onderzoeksmateriaal

Bij alle leerlingen werd een rekentoets, een spellingtoets en vragenlijsten over hun 'zelfbeeld' als spellers en rekenaars[1] afgenomen. Bij alle opgaven van de reken- en de spellingtoets werd hen gevraagd aan te geven hoe zeker zij waren over de juistheid van hun antwoord.

Rekentoets

In overleg met groepsleerkrachten werden 54 rekenopgaven geselecteerd uit de Niveau Test Rekenen-Technisch (De Vos, 1995), zo dat ze merendeels niet te gemakkelijk zouden zijn voor leerlingen van Groep 7 en merendeels vermoedelijk ook geschikt voor leerlingen van Groep 6. De meeste opgaven moesten met hoofdrekenen worden opgelost. Enkele opgaven vroegen om 'cijferen' (sommen 'onder elkaar' oplossen). Tabel 1 geeft een overzicht – met voorbeelden – van de samenstelling.

Tabel 1 De samenstelling van de Rekentoets

	Hoofdrekenen		Cijferen	
Bewerking:	**Aantal**	**Voorbeelden**	**Aantal**	**Voorbeelden**
Optellen	12	*3+2= ….* *68+2500= ….*	3	*Zet onder elkaar en reken uit:* *1103+854+3+267= ….*
Aftrekken	12	*9-6= ….* *3480-1450= ….*	3	*Zet onder elkaar en reken uit:* *8104-4081= ….*
Vermenigvuldigen	8	*9x8= ….* *2x43= ….*	2	*1024* *607 x* *…….*
Delen	6	*35:5= ….* *25:6= …. rest: …*	2	*98:7= ….* *689:13= ….*
Breuken	6	*1= ….../3* *…. deel van 16=4*		

De sommen stonden onder elkaar op een toetsformulier, met achter de sommen twee kolommen. Boven de eerste kolom stond "Ik denk dat mijn antwoord GOED is", boven de tweede "Ik denk dat mijn antwoord FOUT is". De leerlingen kregen een half uur de tijd om de sommen te maken. Ze werden geïnstrueerd na elke som onmiddellijk met een kruis in de betreffende kolom aan te geven of ze dachten dat hun antwoord goed of fout was. Per opgave werd vastgesteld of het antwoord goed (1) of fout (0) was. De itemscores werden opgeteld tot een toetsscore, de *rekenscore*. De itemzekerheidsscores (goed=1, fout=0) werden opgeteld tot *de rekenzekerheidsscore*.

Spellingtoets

De Spellingtoets werd samengesteld met opgaven uit het PI-dictee (Geelhoed & Reitsma, 1999), de woorden die bedoeld zijn voor eind Groep 5, 6 resp. 7. Het betreft de volgende woorden:

- Groep 5: huwelijk, beweging, vergissing, vrachtschepen, ademhaling, draaierig, schattig, zonnetje, voorzichtig, kanonnen, nieuwsgierig, verzameling, kostbare, vreselijk, middelen
- Groep 6: kwaadheid, komma's, bagage, kilometer, tropisch, rivaliteit, vakantie, vrolijkheid, hardste, informatie, verlegenheid, politie, gigantisch, etalage, pagina's
- Groep 7: chirurg, niveau, citroen, vertrouwelijk, stommiteit, centrale, Afrikaanse, ontroerend, hoofdzakelijk, december, explosie, reparatie, discotheek, pannenkoek, ontzaglijk

Elk woord werd gedicteerd in het kader van een zin. De proefleider las eerst die zin in zijn geheel voor en herhaalde daarna het doelwoord. De leer-

lingen schreven de woorden op een formulier met genummerde regels. Achter de ruimte waarop de woorden (en eventuele verbeteringen) konden worden opgeschreven, bevonden zich twee kolommen, met boven de ene kolom "Ik denk dat ik dit woord GOED heb geschreven" en boven de andere "Ik denk dat ik dit woord FOUT heb geschreven". Onmiddellijk nadat ze een woord geschreven hadden, gaven de leerlingen met een kruis hun keus aan in de betreffende kolom. De itemscores werden opgeteld, tot de *spellingscore*, en de itemzekerheidsscores tot de *spellingzekerheidsscore*.

Procedure

De toetsen werden klassikaal ingevuld, uit praktische overwegingen in een vaste volgorde: Spellingtoets (alle leerlingen van de klas schreven tegelijk hetzelfde woord), Rekentoets (klassikaal, maar in het eigen tempo van de leerling). De toetsafname stond onder leiding van de twee student-onderzoekers. De instructies stonden op schrift en werden door de testleiders klassikaal voorgelezen. Elk onderdeel werd met enkele voorbeelditems geoefend. De leerlingen kregen geen feedback over hun scores. Hen werd verteld dat de toetsuitslagen wel aan de leerkrachten zouden worden doorgegeven, als die daar belangstelling voor hadden.

Analyses

Met vier afzonderlijke GLM's wordt het patroon in de scores op de beide schoolvakken en dat in de respectieve zekerheidsscores getoetst. De reken- en spellingscores resp. de zekerheidsscores voor beide vakken zijn daarbij de afhankelijke variabelen, jaargroep en geslacht de factoren. Met een GLM op de verschilscores tussen de proportionele toets- en zekerheidsscores worden de prestatie- en zekerheidpatronen voor beide schoolvakken met elkaar vergeleken. De factoren zijn schoolvak (binnen proefpersonen) en jaargroep en geslacht (tussen proefpersonen). Met overeenkomstige analyses worden terechte acceptatie en verwerping onderzocht.

Resultaten

Het doel van dit onderzoek was te verkennen hoe de zekerheid van leerlingen van Groep 6 en 7 over de kwaliteit van hun antwoorden zich verhoudt tot hun feitelijke kwaliteit. Voor dat doel bepalen we eerst de scores van jongens en meisjes in de Groepen 6 en 7 op toetsen voor spelling- en rekenvaardigheid. Vervolgens stellen we vast of deze vaardigheden en eventuele verschillen weerspiegeld worden in de zekerheid over de gemaakte opgaven.

Reken- en spellingvaardigheid

Het is te verwachten dat Groep 7 beter zal presteren op de *Rekentoets* dan Groep 6. De literatuur die we raadpleegden voor de Inleiding laat ook zien dat de jongens vrijwel zeker beter zullen rekenen dan de meisjes. De gemiddelden voor rekenen in Tabel 2 stemmen overeen met deze verwachtingen. Een GLM-analyse met de *rekenscore* als afhankelijke variabele en met *groep* en *geslacht* als factoren ondersteunt het effect van de *groep* ($F(1, 162) = 6.28$, $p < .05$) en ondersteunt ook de verwachting dat er een effect is van het *geslacht* ($F(1, 162) = 2.72$, $p(eenzijdig) = .05$). De interactie van *groep* en *geslacht* is onbeduidend ($F(1, 162) = 2.64$, $p = .11$).

Tabel 2 Scores op de Rekentoets en de Spellingtoets (gemiddelde en *SD*)

	Rekenen			Spellen		
	Groep 6	Groep 7	Totaal	Groep 6	Groep 7	Totaal
Meisjes	44.93	47.57	46.12	30.51	36.41	33.17
	(4.03)	(4.03)	(4.22)	(7.05)	(4.01)	(6.55)
Jongens	47.02	47.58	47.26	31.27	35.53	33.10
	(3.65)	(4.69)	(4.11)	(6.83)	(5.62)	(6.65)
Totaal	46.01	47.58	46.70	30.90	35.97	33.13
	(3.96)	(4.34)	(4.19)	(6.91)	(4.86)	(6.58)

Deze resultaten zijn ongetwijfeld beïnvloed door de gemakkelijkheid van de Rekentoets en de 'restriction of range' die daarvan het gevolg zal zijn geweest: Gemiddeld maakte men 87% van de sommen goed. Iets minder dan eenvijfde van de leerlingen maakte 50 of meer van de 54 sommen goed. Er is weliswaar een effect van *Groep*, maar dat ene jaar onderwijs heeft geleid tot een scoreverschil van niet meer dan anderhalve som. En datzelfde plafond-effect kan ertoe hebben geleid dat de meisjes de jongens zijn gaan evenaren in Groep 7, in plaats van hun achterstand te behouden of te vergroten. Met deze beperktheid van de toets zullen we in het volgende rekening moeten houden.

De resultaten van de *Spellingtoets* (Cronbach's Alpha: .88) zullen, bij de gemiddelde score van 74% goed, minder aan plafond-effecten onderhevig zijn. Tabel 2 laat dan ook een stevig effect van *groep* zien, dat significant is ($F(1, 162) = 28{,}13$, $p < .001$). Dat het effect van *geslacht*, dat doorgaans gevonden wordt, er niet is ($F < 1$), ook niet in de vorm van een interactie met *groep* ($F < 1$), zal dus niet aan een plafond-effect kunnen worden toegeschreven.

De correlatie tussen de scores op de twee toetsen is relatief gering, .44. Dat komt niet omdat de toetsen te onbetrouwbaar zouden zijn: Cronbach's Alpha is .74 voor de Rekentoets en .88 voor de Spellingtoets. Het bevestigt

veeleer dat het gaat om twee vaardigheden die zich min of meer onafhankelijk van elkaar kunnen ontwikkelen.

Zekerheid dat de opgaven goed of fout gemaakt zijn

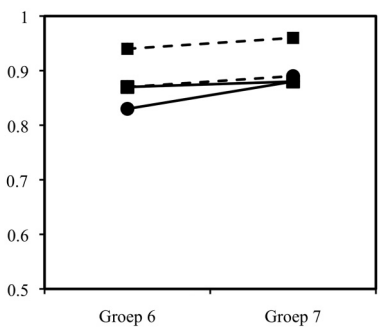
De *zekerheidsscores* voor rekenen en spellen zijn slechts in beperkte mate aan elkaar gerelateerd, .48, ongeveer zoveel als de onderliggende vaardigheden. De zekerheid op beide domeinen zal dus in beperkte mate bepaald worden door een en dezelfde 'algemene zekerheid'. Maar ook de samenhang met de domeinen is beperkt: De correlatie met de *toetsscores* is voor rekenen niet meer dan .47, voor spellen slechts .26.

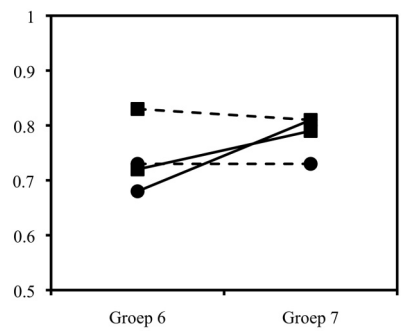
Tabel 3 geeft de gemiddelde zekerheidsscores voor de twee groepen en de beide geslachten.

Tabel 3 Zekerheidsscores voor de Rekentoets en de Spellingtoets (gemiddelde en *SD*)

	Rekenen			Spellen		
	Groep 6	Groep 7	Totaal	Groep 6	Groep 7	Totaal
Meisjes	47.07	48.27	47.61	32.64	32.65	32.65
	(5.04)	(7.66)	(6.33)	(6.63)	(6.99)	(6.75)
Jongens	50.87	51.92	51.32	37.38	36.39	36.95
	(2.58)	(2.57)	(2.61)	(6.09)	(7.02)	(6.48)
Totaal	49.03	50.0	49.49	35.09	34.49	34.83
	(4.38)	(5.90)	(5.16)	(6.75)	(7.21)	(6.94)

De reken- en de spellingtoets verschilden in het aantal items. Om de vergelijking van de toetsen te vergemakkelijken zijn de toetsscores en de zekerheidsscores eerst omgerekend tot proporties van hun respectievelijke maximum (rekenen: 54; spellen: 45). Figuren 1a en 1b illustreren deze gemiddelden.





Figuur 1a Rekenen *Figuur 1b Spellen*

Het aantal goede antwoorden (––) en het aantal oordelen 'zeker goed' (- - -) als proportie van het maximum op de rekentoets (a) en de spellingtoets (b), voor meisjes (●) en jongens (■).

De zekerheid dat de opgave goed gemaakt is blijkt in het algemeen groter dan de feiten rechtvaardigen. Het verschil tussen de (proportionele) *zekerheids-* en *toetsscores* is zowel voor rekenen als voor spellen gemiddeld groter dan 0 (0.05, $t(165) = 7.38$, $p < .001$) resp. 0.04, $t(165) = 2.65$, $p < .01$)). Deze verschilscores zijn significant, maar hangen niet volledig met elkaar samen: $r = .45$.

Nemen we de twee verschilscores als afhankelijke variabelen in een GLM met groep en geslacht als tussen-proefpersonen- en schoolvak als binnen-proefpersonen-factor, dan blijkt er een substantiële interactie te zijn van *groep* en *schoolvak* ($F(1, 162) = 22.91$, $p < .001$), die laat zien dat de gemiddelde discrepantie voor rekenen ongeveer .05 blijft, terwijl de aanvankelijke overschatting voor spellen (.09) zelfs plaats maakt voor een onderschatting (-.03). De meisjes overschatten hun prestaties minder dan de jongens ($F(1, 162) = 18.12$, $p < .001$). De gemiddelde overschatting door de meisjes is vrijwel nihil (.004), maar voor de jongens .08. Er is echter ook een (marginale) interactie van *geslacht* en *schoolvak* ($F(1, 162) = 3.49$, $p = .06$), die laat zien dat de meisjes hun rekenen enigszins overschatten en hun spellen enigszins onderschatten, terwijl de zekerheid van de jongens in beide schoolvakken de feitelijke prestaties overtreft.

Om de discrepantie tussen de feitelijke en de door de leerling zelf beoordeelde kwaliteit van sommen en spellingen te preciseren en de samenhang met groep en geslacht verder te bepalen, is per leerling per opgave vastgesteld of het antwoord goed dan wel fout was en of de leerling het antwoord als goed dan wel fout beschouwde. Tabel 4 geeft per toets de proportionele frequentie van elk van deze vier combinaties.

Tabel 4 De proportionele frequentie van de combinaties van feitelijke en beoordeelde kwaliteit van de antwoorden

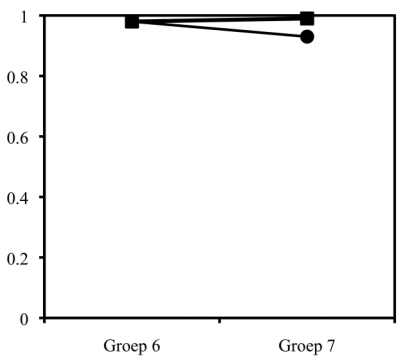
	Feitelijk					
	Rekenen			Spellen		
Beoordeeld:	Goed	Fout	Totaal	Goed	Fout	Totaal
Goed	.84	.08	.92	.64	.13	.77
Fout	.03	.05	.08	.10	.13	.23
Totaal	.87	.13	1	.74	.26	1

We concentreren ons op de correcte oordelen, dus op de onderkenning van goede antwoorden als goed (*'terechte acceptatie'*) en foute antwoorden als fout (*'terechte verwerping'*). De feitelijke vulling van deze cellen is significant hoger dan de proportie die men zou verwachten als alleen de toetsscores en de zekerheidsscores (de 'randtotalen') daarvoor bepalend waren

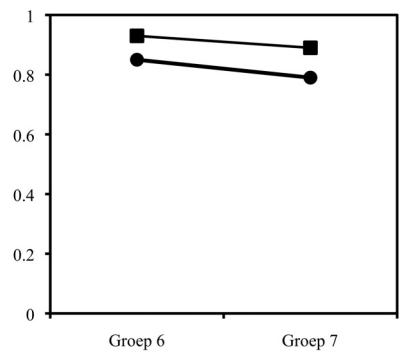
(Rekenen: $t(165) = 12.18$, $p < .001$; Spellen: $t(165) = 26.23$, $p < .001$)[2]. Het is dus kennelijk niet een zaak van louter toeval of leerlingen goed gemaakte opgaven als ´goed´ beoordelen en fout gemaakte als 'fout'; zij hebben er wel enigszins kijk op.

Dat er bij rekenen meer goed gemaakte opgaven als goed werden beoordeeld dan bij spellen, zal te maken hebben met het feit dat er ook meer rekenopgaven goed gemaakt werden. Vergelijkbaars geldt voor de onderkenning van fout gemaakte opgaven als fout. Om rekenen en spellen beter met elkaar te vergelijken, corrigeren we daarom voor de frequentie van de goede resp. foute reken- resp. spellingopgaven. Figuren 2 (*'terechte acceptatie'*) en 3 (*'terechte verwerping'*) illustreren de uitkomsten.

Er vallen een paar zaken op. *Terechte acceptatie* komt vaker voor bij de Rekentoets dan bij de Spellingtoets (.97 vs. .87: $F(1,162) = 136.79$, $p < .001$). Er is een interactie van *groep* en *toets* ($F(1, 162) = 5.75$, $p < .05$): bij het spellen is men in Groep 6 zekerder over de goede antwoorden dan in Groep 7 (.89 vs. .84), terwijl dit groepsverschil bij het rekenen (.97 vs. .96) klein is (zie Figuur 2). Jongens zijn in het algemeen zekerder over hun goede antwoorden dan meisjes (.95 vs. .88): $F(1,162) = 21.11$, $p < .001$. Bij spellen is dat verschil tussen jongens en meisjes (.91 vs. .82) groter ($F(1,162) = 5.13$, $p < .05$) dan bij rekenen (.99 vs. 94).



Figuur 2a Rekenen

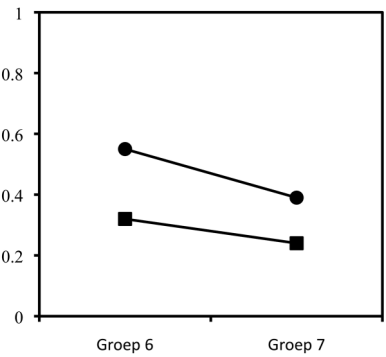


Figuur 2b Spellen

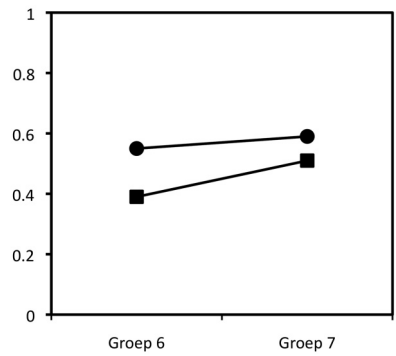
De proportie als 'goed' onderkende goede antwoorden ('terechte acceptatie'), bij rekenen (a) en spellen (b), voor meisjes (●) en jongens (■).

Foute antwoorden worden minder vaak als fout onderkend dan goede antwoorden als goed. Bij spelling komt deze *terechte verwerping* in het algemeen vaker voor ($F(1,160^3) = 25.96$, $p < .05$) dan bij rekenen (.51 vs. .37). Het verschil tussen de toetsen neemt toe van Groep 6 naar Groep 7, omdat de frequentie van terechte verwerping over de sommen daalt (van

.43 naar .32) en die over de spellingen toeneemt (van .47 naar .55): $F(1,160) = 12.02$, $p < .001$ (zie Figuur 3). *Terechte verwerping* komt bij de meisjes vaker voor dan bij de jongens ($F(1,160) = 22.78$, $p < .05$), ongeacht de *groep* ($F(1,160) = 1.10$, ns) en de *toets* ($F(1,160) = 1.13$, ns).



Figuur 3a Rekenen



Figuur 3b Spellen

De proportie als 'fout' onderkende foute antwoorden ('terechte verwerping'), bij rekenen (a) en spellen (b), voor meisjes (●) en jongens (■).

Samenvattend kan men vaststellen dat terechte acceptatie en verwerping niet twee zijden zijn van dezelfde medaille. Goede antwoorden worden vaker als goed onderkend dan foute antwoorden als fout. Deze correcte vaststellingen zijn bij het rekenen anders dan bij het spellen en ontwikkelen zich anders van groep 6 naar 7. Leerlingen gaan meer goede spellingen verwerpen en meer foute spellingen als fout onderkennen, terwijl ze even zeker blijven over goed gemaakte sommen en meer fout gemaakte sommen als goed gemaakt accepteren. De jongens scoren hoger dan de meisjes bij de identificatie van de goede antwoorden en de meisjes hoger bij de identificatie van de foute antwoorden. Dat de correlatie tussen terechte acceptatie en terechte verwerping beperkt en zelfs negatief is (spellen: -.57, rekenen: -.30) wijst erop dat het niet twee aspecten zijn van min of meer dezelfde expertise: weten wat goed is en daardoor ook weten wat fout is.

Discussie

De centrale vraag is in hoeverre bij basisschoolleerlingen (hier beperkt tot leerlingen van de Groepen 6 en 7) verschillen in reken- en spellingvaardig-

heid samengaan met verschillen in zekerheid over hun antwoorden op afzonderlijke opgaven.

De correlatie tussen spelling- en rekenscores past bij wat men in de literatuur aantreft (zie Roebers et al., 2012). Die is, hoewel beduidend, toch zo laag dat het legitiem is in de analyses de beide schoolvaardigheden te onderscheiden.

Verschillen tussen leerlingen in spelling- en rekenvaardigheid hangen samen met het geslacht en de jaargroep van de leerlingen. De leerlingen van Groep 7 bleken uiteraard beter te rekenen dan die van Groep 6. Over het geheel genomen rekenden de jongens beter dan de meisjes, wat in overeenstemming is met veel gepubliceerd onderzoek (o.a. Van Boxtel et al., 2011). Dat het verschil dat in Groep 6 nog duidelijk was, in Groep 7 vrijwel is verdwenen, is wellicht een vertekening die het gevolg is van een plafond in de scoreverdeling. Ook in spellen waren de leerlingen van Groep 7 vanzelfsprekend beter dan die van Groep 6, maar wij vonden niet zoals sommige anderen dat de meisjes daarin beter presteerden dan de jongens. Dat laatste komt misschien doordat dat verschil doorgaans (bijv. Helmke & Van Aken, 1995; Hemker et al., 2010; Van Boxtel et al., 2011) niet groot is, zodat de kans op het vinden van een nul-effect als het onze reëel is.

Ook uit ons onderzoek blijkt, zoals uit dat van Lundeberg et al. (1994), Krebs en Roebers (2012) en Roebers (2002), dat de zekerheid dat de antwoorden goed zijn in het algemeen groter is dan de toetsscores rechtvaardigen. Deze overschatting is echter niet algemeen, want we zien daarin verschillen die samenhangen met het schoolvak (rekenen of spellen), het geslacht en de jaargroep. Het meest interessant is wellicht de interactie van groep en schoolvak die anders is dan wij verwachtten. In plaats van een geringere discrepantie bij rekenen dan bij spellen, die afneemt van Groep 6 naar Groep 7, zien we een complexer beeld: Terwijl de overschatting van de rekenprestaties nauwelijks verandert van de ene jaargroep naar de andere, lijkt de overschatting van de spellingprestaties in Groep 6 te veranderen in een onderschatting in Groep 7.

Hoe kan het dat de lichte stijging van de rekenscores samengaat met een lichte toename van de zekerheid over de prestaties, terwijl de spellingscores flink verbeteren zonder dat dit zijn weerslag heeft in de manier waarop de leerlingen hun werk waarderen en die waardering zelfs relatief achteruit gaat? Onze verklaring is eenvoudig. De leerlingen geven met hun oordeel telkens te kennen of zij denken te beschikken over de kennis en de vaardigheid die naar hun idee nodig waren voor de opdracht. Bij dit inschatten van de vereiste competentie zou het maken van een gegeven som

beschouwd kunnen worden als een overzichtelijke toepassing van enkele regels, waarin men zich steeds meer bekwaamt. De inschatting van wat men moet kunnen om die som te maken verandert niet wezenlijk, wel neemt het vertrouwen in de vereiste bekwaamheid toe.

Het spellen van een gegeven woord wordt misschien aanvankelijk ook beschouwd als niet meer dan de weergave van de opeenvolgende spraakklanken via enkele simpele regels. Maar gaandeweg groeit vermoedelijk de argwaan dat je bij het spellen van een woord nooit zeker weet of het geen uitzondering bevat en of er geen bijzondere regels op van toepassing zijn. Kortom, de interactie laat zien hoe de leerlingen de competentie inschatten die nodig is voor deze woorden en deze sommen; en hoe die opvattingen – voor spellen anders dan voor rekenen – tijdens de schoolloopbaan kunnen veranderen.

Bovendien, we suggereerden het eerder al, kunnen sommen tamelijk eenvoudig gecontroleerd worden op hun juistheid, met name door herhaling van de bewerking of door 'terugwaartse' toepassing (zoals na een aftreksom door optelling van het afgetrokkene bij het resultaat). Bij het spellen zijn herhaald schrijven en teruglezen niet vergelijkbaar probate middelen. Daar is woordspecifieke orthografische kennis (vaak op basis van morfologie of etymologie) nodig, zelfs om te weten of je een gegeven woord rechttoe rechtaan mag schrijven (Henneman, 2000; Kleijnen, 1988; Van Bon, 1993). Terwijl je dus van sommen leert inschatten hoe je ze moet maken (wanneer je er een van een bepaalde soort kunt, kun je ze allemaal), ontdek je dat het kunnen schrijven van het ene woord niet inhoudt dat je vergelijkbaar klinkende woorden ook aankunt. Deze gedachtegang wordt ondersteund door de constatering dat in Groep 7 de terechte acceptatie van goede antwoorden bij het rekenen nauwelijks anders is dan in Groep 6 (wellicht zelfs was toegenomen als de rekenscore niet zijn plafond had bereikt), maar bij het spellen afneemt, terwijl de terechte verwerping van fout gemaakte sommen afneemt en de terechte verwerping van foutieve spellingen juist toeneemt.

Deze uitkomsten en overwegingen verklaren de beperkte samenhang tussen toetsscores en zekerheid, die voor spellen lager is dan voor rekenen. Ze maken ook eens te meer duidelijk dat het leren van de schoolvakken niet over de gehele linie bestaat uit een toename in vaardigheid en kennis die gepaard gaat met toenemend vertrouwen in wat men geleerd heeft. Het lijkt ons van belang deze speculaties over het relevante verschil tussen rekenen en spellen – twee belangrijke basisschoolvakken – te formaliseren en aan toetsing te onderwerpen.

In de inleiding maakten we al melding van onderzoek dat in strijd is

met de opvatting, van bijvoorbeeld Pallier et al. (2002) dat zekerheid een algemeen persoonskenmerk zou zijn dat de beoordeling van alle soorten opgaven in gelijke mate betreft. Ons onderzoek ondersteunt die opvatting evenmin. De zekerheid bij rekenen en die bij spellen zijn immers slechts matig gecorreleerd. Ook is de frequentie van terechte acceptatie en terechte verwerping niet het simpele product van de toets- en zekerheidsscores. Bovendien wijst de negatieve correlatie tussen terechte acceptatie en terechte verwerping erop dat het niet twee aspecten zijn van min of meer dezelfde zekerheid. Het valt dan ook te betwijfelen of de optelling van beide soorten zekerheden (Roebers et al., 2012), bijvoorbeeld tot 'spellingbewustzijn' (Paffen & Bosman, 2005), terecht is. Leerlingen differentiëren en nuanceren kennelijk bij het geven van hun zekerheidsoordeel, naar leerstofdomein, naar eigenschappen van de opgaven binnen een domein en naar het soort oordeel (acceptatie of verwerping). Het mag dan zo zijn dat leerlingen van elkaar verschillen in de zekerheid die ze meestal vertonen over hun schoolprestaties, hun zekerheid is blijkbaar niet zo algemeen dat hij bij alle schoolvakken en daarbinnen bij alle soorten opgaven dezelfde is. De kijk die leerlingen hebben op hun bekwaamheid is complex en meervoudig.

Ook de suggestie van Kruger en Dunning (1999), dat overschatting neigt om te slaan in onderschatting als de competentie zijn maximum nadert, moet kennelijk worden genuanceerd. Want terwijl we bij spellen de twijfel zien toenemen, is dat bij het rekenen niet te merken. Verschillen in de aard van de kennis en vaardigheden die betrokken zijn bij rekenen en spellen zouden daarvan de oorzaak kunnen zijn, ook een reden om voortaan het concept 'metacognitie' te differentiëren en te onderzoeken in samenhang met de onderhanden taken.

In overeenstemming met het merendeel van de literatuur (o.a. Boekaerts & Rozendaal, 2010; Job & Klassen, 2012) blijkt de zekerheid van jongens groter te zijn dan hun prestaties rechtvaardigen, voor beide schoolvakken. Als hun vertrouwen inzake taal gaandeweg afneemt (Jacobs et al., 2002), dan is dat in Groep 6 en 7 nog niet merkbaar. Gemiddeld zijn de zekerheidsoordelen van de meisjes in evenwicht: ze overschatten hun rekenen enigszins, maar onderschatten hun spellen. De te grote zekerheid van jongens is deels het gevolg van hun grotere vertrouwen in geval van goede antwoorden, maar ook van hun onterechte acceptatie van foute antwoorden. Meisjes blinken daarentegen uit in de identificatie van foute antwoorden. Dat laatste werd ook vastgesteld door Lundeberg et al. (1994, 2000), maar anders dan wij vonden zij in geval van goede antwoorden geen grotere zekerheid bij jongens. Of en onder welke condities er zo'n ge-

slachtsgebonden verschil in de evaluatie van goede en foute antwoorden te vinden is, zal verder onderzoek dus moeten uitwijzen. Het is echter waarschijnlijk dat zekerheidsoordelen – anders dan Kröner en Bierman (2007) en Pieschl (2009) suggereren – wel afgestemd worden op kenmerken van de te vervullen opdrachten. Zij lijken niet alleen te berusten op eerdere ervaring en wel te worden aangepast aan de opgave van het moment. Zij weerspiegelen dus kennelijk zich voltrekkende metacognitieve processen, bij meisjes vermoedelijk meer dan bij jongens.

Naar goed gebruik wijzen we op enkele beperkingen van ons onderzoek.

Op de eerste plaats is daar de keuze van de toetsopgaven. Door het ontbreken van een toegankelijk steekproefkader zijn daar willekeurige selecties voor gebruikt. Het is dus niet zeker of deze woorden en sommen representatief zijn voor de spelling- en rekenleerstof. Replicatie van dit onderzoek met andere toetsopgaven is dus gewenst, wat de gelegenheid zou geven na te gaan of de zekerheidsoordelen variëren met de eigenschappen van de opgaven. Vooral bij spelling wordt het interessant of de beperkte zekerheid die we hier vonden, zich over alle woorden uitstrekt of op een verklaarbare manier samenhangt met bepaalde woordeigenschappen.

Een andere beperking inzake de keus van de opgaven is dat ze zowel door leerlingen van Groep 6 als Groep 7 te maken moesten zijn, wat tot gevolg had dat somtypen die geschikter waren voor Groep 7 niet in aanmerking kwamen en de rekentoets nogal gemakkelijk was. Dat heeft echter het vinden van belangrijke verschillen in en in samenhang met de rekentoets niet in de weg gestaan.

Het is ook niet uitgesloten dat onze leerlingensteekproef een vertekend beeld geeft. Het ging immers om slechts drie scholen en dus om een beperkt aantal leermethoden en pedagogische benaderingen. Zo is het mogelijk dat gangbare verschillen tussen geslachten of Groepen in vaardigheid, zekerheid of zelfbeeld uitgewist zijn door de toevallige onderwijspraktijk op een enkele van dit beperkte aantal scholen.

Deze beperkingen van het onderzoek doen echter niet af aan enkele belangrijke constateringen. De eerste en centrale constatering is dat de feitelijke bekwaamheid, zoals die blijkt uit de toetsscores, slechts in beperkte mate bepalend is voor de zekerheid op het gebied van rekenen en spellen. Maar juist door zijn invaliditeit als indicator van het prestatieniveau kan zekerheid een waardevolle informatiebron zijn over het functioneren van de leerling. De tweede is de bevestiging van het feit dat leerlingen doorgaans (maar jongens veel meer dan meisjes) te zeker zijn over

hun antwoorden, waardoor hun neiging om de opgaven secuur aan te pakken en hun antwoorden te heroverdenken beperkt zal zijn. De derde is dat deze zekerheid geen algemeen persoonlijkheidskenmerk is, maar met het schoolvak en met de bekwaamheid varieert. De vierde constatering is dat het schoolvak *spellen* er voor veel leerlingen aanvankelijk even toegankelijk en betrouwbaar uitziet als rekenen, maar gaandeweg een domein wordt dat met argwaan wordt betreden vanwege de vele adders onder het gras. Het belang van deze constateringen pleit voor replicatie van dit onderzoek over een uitgebreider bereik (naar schoolvakken en groepen) van het basisonderwijs, waarbij het ontwerpen van adequate kaders voor de selectie van representatieve en informatieve opgaven extra aandacht vraagt.

Noten

1. Over het zelfbeeld van de leerlingen wordt in een aparte publicatie verslag gedaan. Eefke van Geffen en Danielle Sabandar werkten de onderzoeksopzet uit en deden de dataverzameling voor de masteropleiding Orthopedagogiek aan de Radboud Universiteit Nijmegen. Wij zijn hen dankbaar voor hun energieke inzet.
2. De proportie goed gemaakte rekenopgaven die als 'goed gemaakt' beoordeeld werden, bijvoorbeeld, bedraagt .84, terwijl men op grond van de marginale frequentie .92*.87 = .80 zou verwachten. Merk op dat per schoolvak het verschil tussen de verwachte en de feitelijke proporties in de cellen gelijk is, zodat de toetsing van het verschil voor alle vier de cellen tot hetzelfde resultaat leidt.
3. Van twee deelnemers konden de gegevens niet in de berekening worden betrokken omdat zij geen fouten maakten en de bepaling van de proportie dus een deling door 0 inhield.

Referenties

Adams, P.A., & Adams, J.K. (1960). Confidence in the recognition and reproduction of words difficult to spell. *American Journal of Psychology*, *73*, 544-552.

Block, K.K., & Peskowitz, N.B. (1990). Metacognition in spelling: Using writing and reading to self-check spellings. *Elementary School Journal*, *91*, 151-164.

Boekaerts, M., & Rozendaal, J.S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction*, *20*, 372-382.

Bonham, A.J., & Gonsalves-Vallejo, C. (2009). Assessment of calibration for reconstructed eyewitness memories. *Acta Psychologica*, *131*, 34-52.

De Vos, T. (1995). *Niveau Test Rekenen – Technisch.* Lisse: Swets.

Desoete, A. (2009). Metacognitive prediction and evaluation skills and mathematical learning in third-grade students. *Educational Research and Evaluation*, *15*, 435-446.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, *117*, 429-456.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 552-564.

Geelhoed, J., & Reitsma, P. (1999). *PI-dictee.* Lisse: Swets & Zeitlinger.

Helmke, A., & Van Aken, M.A.G. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: a longitudinal study. *Journal of Educational Psychology*, *87*, 624-637.

Hemker, B.T., Kuhlemeier, J.B., & Van Weerden, J.J. (2010). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2009.* Arnhem: Cito.

Henneman, K. (2000). *Problemen van gevorderde spellers.* Bussum: Coutinho.

Jacobs, J.E., Lanza, S., Osgood, D.W., Eccles, J.S., & Wigfield, A. (2002). Changes in children's self-competence and values: gender and domain differences across grades one through twelve. *Child Development*, *73*, 509-527.

Job, J., & Klassen, R.M. (2012). Predicting performance on academic and non-academic tasks: A comparison of adolescents with and without learning disabilities. *Contemporary Educational Psychology*, *37*, 162-169.

Jonson, A.-C., & Allwood, C. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personality and Individual Differences*, *34*, 559-574.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: a critical examination of the hard-easy effect. *Psychological Review*, *107*, 384-396.

Kasperski, R., & Katzir, T. (2013). Are confidence ratings test- or trait-driven? Individual differences among high, average, and low comprehenders in fourth grade. *Reading Psychology*, *34*, 59-84.

Kleijnen, R. (1988). *Hardnekkige spellingfouten. Een taalkundige analyse.* Lisse; Swets & Zeitlinger.

Krebs, S.S., & Roebers, C.M. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *British Journal of Educational Psychology*, *40*, 325-340.

Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1121-1134.

Kröner, S., & Bierman, A. (2007). The relationship between confidence and self-concept – Towards a model of response confidence. *Intelligence*, *35*, 580-590.

Lundeberg, M.A., Fox, P.W., & Puncochar, J. (1994). Highly confident but wrong: gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, *86*, 114-121.

Lundeberg, M.A., Fox, P.W., Brown, A.C., & Elbedour, S. (2000). Cultural influences on confidence: Country and gender. *Journal of Educational Psychology*, *92*, 152-159.

Marsh, H.W., & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin*, *34*, 542-552.

Marsh, H.W., Roche, L.A., Pajares, F., & Miller, D. (1997). Item-specific efficacy judgments in mathematical problem solving: The downside of standing too close to trees in a forest. *Contemporary Educational Psychology*, *22*, 363-377.

Paffen, R., & Bosman, A.T.M. (2005). Spellingbewustzijn kan met een korte training gestimuleerd worden. *Tijdschrift voor Orthopedagogiek*, *44*, 388-397.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. (2002). The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology*, *129*, 257-299.

Pieschl, S. (2009). Metacognitive calibration – An extended conceptualization and potential applications. *Metacognition Learning*, *4*, 3-31.

Roebers, C.M. (2002). Confidence judgments in children's and adults' event recall and suggestibility. *Developmental Psychology*, *38*, 1052-1067.

Roebers, C.M., Cimeli, P., Röthlisberger, M., & Neuenschwander, R. (2012). Executive functioning, metacognition, and self-perceived competence in elementary school children: an explorative study on their interrelations and their role for school achievement. *Metacognition and Learning*, *7*, 151-173.

Schneider, W. (2001). Metacognition and memory development in childhood and adolescence. In H.S. Waters & W. Schneider (Eds.), *Metacognition, strategy use and instruction* (pp. 54-81). New York: Guildford Press.

Schraw, G., Dunkle, M.G., Bendixen, L.D., & DeBacker Roedel, T. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology*, *87*, 433-444.

Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology*, *100*, 961-976.

Van Bon, W.H.J. (1993). *Spellingproblemen. Theorie en praktijk*. Rotterdam: Lemniscaat BV.

Van Boxtel, H., Engelen, R., & De Wijs, A. (2011). *Wetenschappelijke verantwoording van de Eindtoets 2010*. Arnhem: Cito.

Over de auteurs

Wim van Bon was tot aan zijn pensioen universitair hoofddocent bij het Behavioural Science Institute en de sectie Orthopedagogiek van de Radboud Universiteit Nijmegen.

Cecile Kuijpers is docent/onderzoeker bij Pedagogische Wetenschappen en Onderwijskunde (sectie Orthopedagogiek en sectie Onderwijskunde) van de Radboud Universiteit Nijmegen.
E-mail: c.kuijpers@pwo.ru.nl